

Evaluation of the use of psychometric scales in human–wildlife interaction research to determine attitudes and tolerance toward wildlife

Whitehouse-Tedd, K., Abell, J. & Dunn, A.

Published PDF deposited in Coventry University's Repository

Original citation:

Whitehouse-Tedd, K, Abell, J & Dunn, A 2021, 'Evaluation of the use of psychometric scales in human–wildlife interaction research to determine attitudes and tolerance toward wildlife', *Conservation Biology*, vol. 35, no. 2, pp. 533-547.

<https://dx.doi.org/10.1111/cobi.13599>

DOI 10.1111/cobi.13599

ISSN 0888-8892

ESSN 1523-1739

Publisher: Wiley

© 2020 The Authors. *Conservation Biology* published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

Evaluation of the use of psychometric scales in human–wildlife interaction research to determine attitudes and tolerance toward wildlife

K. Whitehouse-Tedd ¹, J. Abell ², and A. K. Dunn ³

¹School of Animal, Rural and Environmental Sciences, Nottingham Trent University Brackenhurst Lane, Southwell, Nottinghamshire, NG25 0QF, U.K.

²Research Centre for Agroecology, Water & Resilience, Coventry University, Ryton Organic Gardens, Coventry, West Midlands, CV8 3LG, U.K.

³School of Social Sciences, Nottingham Trent University, Goldsmith Street, Nottingham, NG1 4BU, U.K.

Abstract: Studies evaluating human–wildlife interactions (HWIs) in a conservation context often include psychometric scales to measure attitudes and tolerance toward wildlife. However, data quality is at risk when such scales are used without appropriate validation or reliability testing, potentially leading to erroneous interpretation or application of findings. We used 2 online databases (ProQuest Psych Info and Web of Science) to identify published HWI studies that included attitude and tolerance. We analyzed these studies to determine the methods used to measure attitudes or tolerance toward predators and other wildlife; determine the proportion of these methods applying psychometric scales; and evaluate the rigor with which the scales were used by examining whether the psychometric properties of validity and reliability were reported. From 2007 to 2017, 114 published studies were identified. Ninety-four (82%) used questionnaires and many of these (53 [56%]) utilized a psychometric scale. Most scales (39 [74%]) had at least 1 test of reliability reported, but reliance on a single test was notable, contrary to recommended practice. Fewer studies (35 [66%]) reported a test of validity, but this was primarily restricted to structural validity rather than more comprehensive testing. Encouragingly, HWI investigators increasingly utilized the necessary psychometric tools for designing and analyzing questionnaire data, but failure to assess the validity or reliability of psychometric scales used in over one-third of published HWI attitude research warrants attention. We advocate incorporation of more robust application of psychometric scales to advance understanding of stakeholder attitudes as they relate to HWI.

Keywords: coexistence, conflict, predator, psychological research methods, reliability, validity, value

Análisis del Uso de Escalas Psicométricas en la Investigación sobre la Interacción Humano-Fauna para Determinar Actitudes y Tolerancia hacia la Fauna

Resumen: Los estudios que analizan las interacciones humano-fauna (IHF) dentro de un contexto de conservación con frecuencia incluyen escalas psicométricas para medir las actitudes y la tolerancia hacia la fauna. Sin embargo, la calidad de los datos se encuentra en riesgo cuando dichas escalas se usan sin una validación apropiada o una prueba de confiabilidad, lo que potencialmente puede llevar a interpretaciones o aplicaciones erróneas de los resultados. Usamos dos bases de datos virtuales (ProQuest Psych Info y Web of Science) para identificar estudios publicados sobre las IHF que incluyeran actitud y tolerancia. Analizamos estos estudios para determinar los métodos utilizados para medir las actitudes o la tolerancia hacia los depredadores y otros tipos de fauna; determinar la proporción de estos métodos aplicando escalas psicométricas; y evaluar el rigor con

email katherine.whitehousetedd@ntu.ac.uk

Article impact statement: Understanding attitudes toward wildlife in human–wildlife interaction research requires robust application of social science methods.

Paper submitted December 18, 2018; revised manuscript accepted July 21, 2020.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

el cual se usaron las escalas al examinar si las propiedades psicométricas de validez y confiabilidad estuvieron reportadas en el estudio. Identificamos 114 estudios publicados entre 2007 y 2017. De estos estudios, 94 (82%) usaron cuestionarios y muchos de estos cuestionarios (53 [56%]) usaron una escala psicométrica. La mayoría de las escalas (39 [74%]) tuvieron al menos una prueba de confiabilidad reportada, pero la dependencia de una sola prueba fue notable, contrario a la práctica recomendada. Fueron menos los estudios (35 [66%]) que reportaron una prueba de validez, pero esto estuvo restringido primordialmente a una validez estructurada en lugar de un análisis más integral. De manera alentadora, los investigadores de las IHF cada vez usaron más las herramientas psicométricas necesarias para diseñar y analizar los datos de los cuestionarios, aunque la falta de análisis de la validez o confiabilidad de las escalas psicométricas utilizadas en más de un tercio de los estudios publicados sobre las actitudes hacia las IHF requiere de atención. Promovemos la incorporación de una aplicación más sólida de las escalas psicométricas para propiciar el entendimiento de las actitudes de los actores sociales conforme se relacionan con las IHF.

Palabras Clave: coexistencia, confiabilidad, conflicto, depredador, métodos de investigación psicológica, validez, valor

【摘要】: 在保护背景下评估人类与野生动物相互作用的研究常常会使用心理测量量表来测量人们对野生动物的态度和容忍度。然而, 如果在使用这些量表时没有进行适当的验证或可靠性测试, 则数据质量可能会存在问题, 还有可能导致对结果的错误解释或应用。本研究使用两个在线数据库 (ProQuest Psych Info和Web of Science) 确定 已发表的包含态度和容忍度的人类与野生动物互作研究。我们通过分析这些研究, 确定 其中衡量人们对食肉动物和其它野生动物的态度或容忍度的方法, 以及这些方法应用心理测量量表的比例, 并通过检查研究是否报告 心理计量学的有效性和可靠性, 对量表使用的严谨性进行 评估。我们分析 2007 年到 2017 年间 114 项已发表的研究, 其中有 94 篇 (82%) 使用 问卷调查, 且有很多 (53篇, 占56%) 使用 心理测量量表。大多数使用量表的研究 (39篇, 占74%) 至少报告一个 可靠性测试, 但需要注意的是仅依赖单一测试并不符合推荐的做法。 有较少研究 (35篇, 占66%) 报告 有效性测试, 但主要局限于结构有效性, 而不是更全面的测试。令人鼓舞的是, 人类与野生动物互作的调查者正在越来越多地使用必要的心理测量工具来设计和分析问卷数据, 但仍有超过三分之一已发表的相关态度研究 能评估心理测量量表的有效性或可靠性, 这需要引起关注。 后, 我们建议使用更稳健的心理测量量表, 以增进对与人类与野生动物互作的利益相关者态度的理解。**【翻译: 胡怡思; 审校: 聂永刚】**

关键词: 共存, 冲突, 心理学研究方法, 食肉动物, 可靠性, 有效性, 价值

Introduction

Human-wildlife interactions (HWIs) occur across the globe and can be defined as events involving direct or indirect contact between humans and nondomestic species (either as individuals, groups, or populations). These HWI on agricultural lands are typically undesirable with severe and negative outcomes, representing a significant threat to many wild species (Dickman 2010) and potentially leading to financial losses or reduced quality of life for people (Baker et al. 2008). As such, a biosocial approach, incorporating human psychology with wildlife biology (e.g., Perry et al. 2020), is pertinent to wildlife conservation research because human dimensions are central to both the problem and solution (Moon & Blackman 2014; Martin 2020).

Studies of the human dimensions of HWI in conservation typically involve measuring stakeholder attitudes toward wildlife on the premise that attitudes are important factors underpinning tolerance and behavioral responses to wildlife (Decker et al. 2012; Delibes-Mateos 2014; Dietsch et al. 2017). Changing attitudes or tolerance toward wildlife are often central aims in mitigating negative HWIs (e.g., Kansky & Knight 2014; Dietsch et al. 2017; Pooley et al. 2017).

However, attitudes are complex intangible constructs with variable definitions and means of quantification or characterization. Broadly, attitudes can be defined as a relatively enduring system of beliefs, feelings, and behavioral tendencies toward something (Hogg & Vaughan 2005). More specifically, attitudes have been defined as a human intention or psychological tendency toward an entity that arises following a positive or negative evaluation of that entity (Ajzen & Fishbein 1980; Eagly & Chaiken 1993; Eagly et al. 1994). Attitudes provide meaning, self-expression, and identity, as well as facilitating social acceptance and protecting self-esteem (Katz 1960). Within social psychology, the ABC model is one means by which responses to attitudinal objects can be understood, or by which the structure of attitudes can be defined. In this model, attitude responses include affect (i.e., emotions), behavior (i.e., verbal and nonverbal), and cognition (i.e., knowledge and beliefs). However, the relative importance of each in determining the attitude is debated and complicated by the synergy between them (Eagly et al. 1994). Moreover, the relationship between attitudes and behavior is not necessarily straightforward, predictive, or consistent (LaPiere 1934; Ajzen et al. 1982; Nilsson et al. 2020). The factors involved in shaping attitudes and responses to attitudinal objects are explored

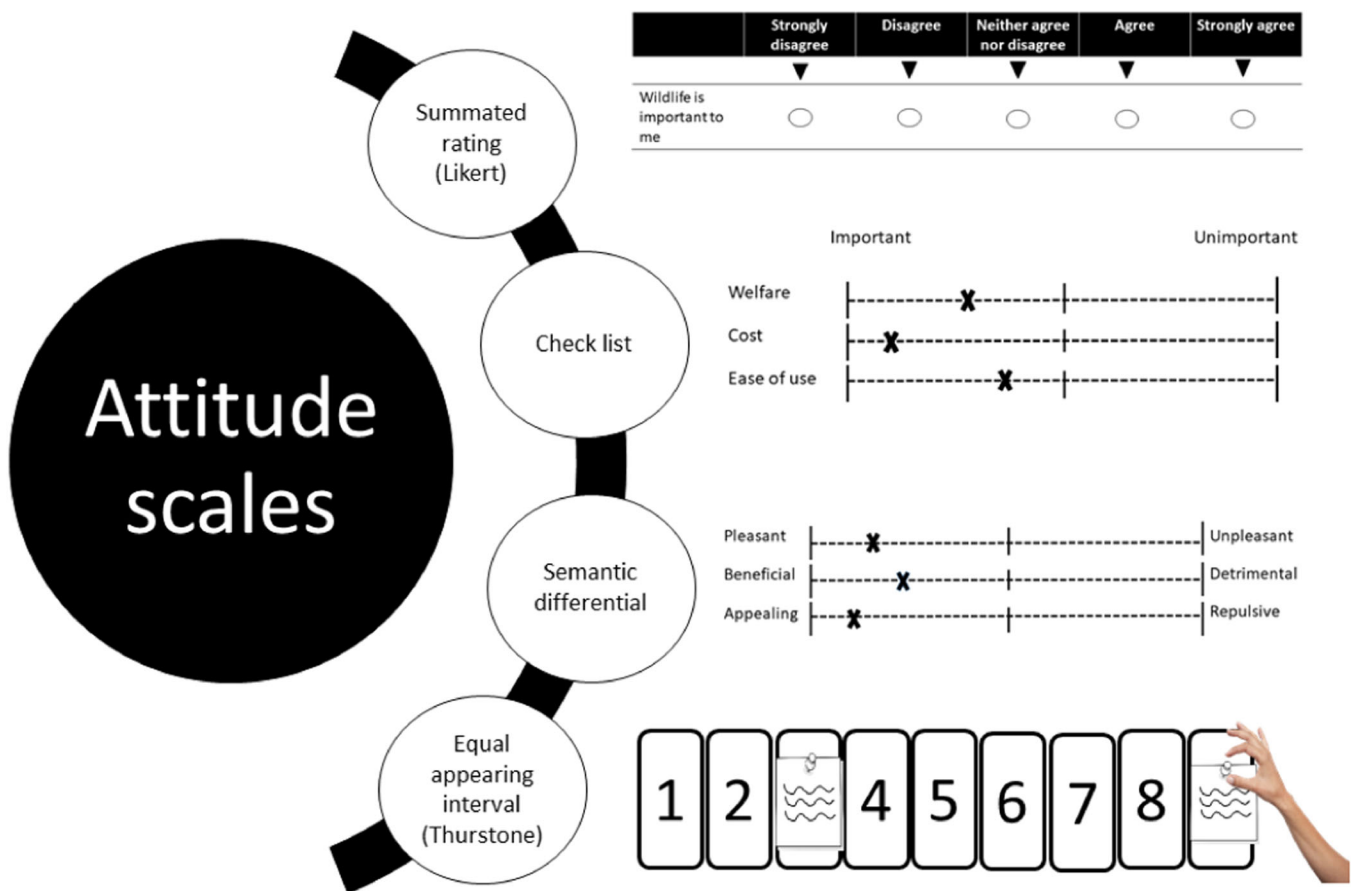


Figure 1. Attitude scales used in human-wildlife interaction studies.

elsewhere (Fulton et al. 1996; Vaske & Donnelly 1999; Vaske & Whittaker 2004), but for attitude research to contribute meaningfully to understanding human behavior in the context of HWI, attitude specificity must be acknowledged on the action, target, context, and time junctures (Ajzen 1985; Manfredo & Bright 2008; Dietsch et al. 2017). In the case of HWI, this could include attitudes toward the setting of traps (action) and capturing wildlife (target) during working hours (time) or on a specific property (context), therefore, making generalization or extrapolation challenging.

Tolerance or intolerance of wildlife can be considered an extension of a person's attitude, either as a means of judging their perception of wildlife, as a construct likely reflective of their behavior (Treves & Bruskotter 2014), or combination of both attitude and behavior. However, erroneous and interchangeable use of the terms *attitude* and *behavior*; assuming a change in attitude will be reflected in behavior; or using generalized attitudes as a proxy for or predictor of highly specific conservation behaviors can result in misdirected conservation efforts, especially those focused on conservation education (Heberlein 2012; Nilsson et al. 2020).

Measuring attitudes is challenging because they cannot be observed directly. Instead, one must observe par-

ticipant's responses toward an object that are considered reflective of their attitude. Concerned with the theory and techniques of psychological measurement, the field of psychometrics has seen the development of a range of scales used to assess attitudes or tolerance. These scales use multiple questions or statements (known as items) for which respondents rate their agreement or alignment with, so the research can derive data for a focal variable (e.g., attitude). As such, scales provide a quantitative approach to understanding human attitudes. Figure 1 contains some scale types commonly used in HWI research. The scales are described in Appendix S1.

Attitude Scales

A diverse range of scales exist to measure psychological constructs. Scales commonly applied to the measurement of attitudes, including in HWI research, range from summated rating (Likert-type) scales, to check-list and semantic differential scales (both of which may utilize a visual analogue scale for response measurement), to a type of Thurstone scale called equal appearing interval scales (Fig. 1 & Appendix S1).

Three key factors must be considered when using attitude scales—reliability, validity, and standardization (Langdridge & Hagger-Johnson 2013a; Coolican 2014)—and their application in wildlife management has been reviewed (Decker et al. 2012). The results of an attitude measure should be tested for reliability (Table 1), such as consistency across time frames (considered stable, or externally consistent), and each item should produce results consistent and correlated with those of other items (internal consistency) (Langdridge & Hagger-Johnson 2013a; Coolican 2014). Additionally, researchers should check for scale validity (i.e., how closely the network of correlations support a construct) (Langdridge & Hagger-Johnson 2013a) (Table 2). Standardization can refer either to the procedure, the interpretation, or the scores generated from psychometric scales (Fischer & Milfont 2010). The standardization of procedures is already a familiar concept in the natural sciences, whereby a form of experimental control is included (Fischer & Milfont 2010). Interpretation standardization typically requires the use of an appropriate social norm with which to compare raw scores (e.g., an average performance of the test when conducted in a representative sample of the population of interest) (Fischer & Milfont 2010). Finally, standardization of scores links to interpretation standardization in that raw scores are expressed relative to the population average and are often called *z* scores (Fischer & Milfont 2010). Standardization is crucial for multisite or interstudy comparisons so that scores can be interpreted consistently within the population or between samples (Coolican 2014). Yet, it must also be recognized that explicit attitudes are not entirely stable. The accuracy and durability of people's responses can vary, especially when understanding, knowledge, or experience of a topic are limited (Manfredo & Bright 2008) or change over time (Nilsson et al. 2020). As such, contextualization may be a higher priority than standardization in some cases.

In light of the criteria for testing reliability and validity, it could be assumed that the best approach to measuring attitudes is to utilize an existing, prevalidated, and reliable scale (Shelby 2011). However, validation of any scale is likely to include some degree of sample specificity. Accordingly, it must have its continued validity for a different sample confirmed (Chase 2016). This is especially important when respondents speak a language different from that in which the scale was originally constructed and validated (thereby requiring validation of the translation process) or when species have multiple common names or even no name in local dialects or languages (e.g., Dickman 2008; Stevens et al. 2014). Solutions to these challenges include the use of symbols or photos (Maddox 2003; Dickman 2008; Hartel et al. 2015), but this often requires concurrent assessment of species identification skills in the respondents when wildlife are the subject of investigation.

It is apparent that no scale is perfect and there are important considerations for their use, which necessitates careful selection and utilization procedures during the study-design phase. Investigations of attitudes conducted without an appreciation of a scale's psychometric properties, limitations, or functions (i.e., without adopting relevant and necessary methodological rigor) can generate misleading conclusions (Moon & Blackman 2014; Martin 2020). However, evidence of poor social science practice has been identified in the published conservation literature (reported by Heberlein 1988; Moon & Blackman 2014; Martin 2020) and inadvertently perpetuates the use of inappropriate methods or interpretation. Not only does this impede the advancement of knowledge, but it also reduces the ability to find effective solutions to environmental problems (Moon & Blackman 2014; Martin 2020). Concerns have been raised regarding the tendency of natural scientists to undertake social science investigations without sufficient knowledge, training, or experience (e.g., Heberlein 1988; Rust et al. 2017; Martin 2020). Fundamentally, authors from outside the social sciences typically ignore prerequisite literature in social science theories and methods. This results in the misapplication of the appropriate methods, which thereby produces unreliable and potentially invalid data, as well as often a failure to report essential methodological information (Heberlein 1988, 2012; Moon & Blackman 2014; Martin 2020).

Although conservation attention to human behavior change needs refocusing (Nilsson et al. 2020), attitudinal change (and therefore measurement) remains among the suite of factors required to facilitate long-term behavior change. Given this, in combination with the concerns raised regarding psychometric scale usage in conservation science, we explored the methods used in HWI studies to measure attitudes or tolerance (hereafter referred to as attitude collectively) toward wildlife and highlight areas for improvement in current practice. We evaluated the use of questionnaires, fixed and semistructured interviews, and focus groups as they relate to HWI. Our analysis of these methods relates to the broader topic of investigating attitudes and tolerance as applied to HWI. Finally, we evaluated the specific use of psychometric scales for measuring attitudes or tolerance in HWI research, including reliability and validity testing and the identification and management of bias.

Literature Search

Socially or ecologically important HWIs in agricultural settings include a diverse range of species. However, of all taxa involved, predators are particularly well represented (Seoraj-Pillai & Pillay 2017). Rather than being of relevance to a relatively select group of species within a taxon (e.g., elephants within the megaherbivores), some

Table 1. Measures of reliability of psychometric scale relevant to human–wildlife interaction studies.

<i>Reliability measure</i>	<i>Test</i>	<i>Description</i>
External consistency	test-retest reliability	Of particular relevance for constructs that are considered to have high stability, test-retest reliability determines external consistency by repeating the use of the scale with the same people across 2 or more different time points; correlation analysis is used to detect random error (Furr 2011a; Langdridge & Hagger-Johnson 2013a; Coolican 2014). Between 2 and 4 weeks is considered appropriate to reduce the potential for priming or practice effect (Furr 2011a, 2011b; Langdridge & Hagger-Johnson 2013a), while also minimizing the risk that an unknown factor has influenced the respondent's attitude such that a change over time is real. However, if a consistent change (systematic error) occurs in the way respondents answer each item between the time of first and subsequent retest, the test is still considered reliable (Rust & Golombok 2009a). This systematic error may be an important factor to understand in itself. In these situations, intraclass coefficient analysis (a technique that determines the similarity of items to each other within a group) can be a useful method for detecting systematic error, as well as external consistency (Furr 2011a).
	inter-rater reliability	When different interviewers conduct interviews or are involved in the coding of data obtained during interviews, it is necessary to test inter-rater reliability (Rust & Golombok 2009a; Furr 2011a). Intraclass coefficient analysis is useful for testing inter-rater reliability and is more flexible than other methods, such as Cohen's kappa, which takes into account agreement between scores occurring by chance (Furr 2011a).
	control group	In scenarios involving a conflict mitigation strategy, a test-retest should be conducted during the before phase (see above). However, the use of a control group in both phases would be necessary to test for external consistency. This design is ethically challenging in human–wildlife interaction (HWI) studies given that withholding interventions that could be either life saving or significantly improve well-being from a subsample of the study population may be met with significant resistance.
Internal consistency	split-half method	This test involves testing the correlation between scores generated by a random sample of half the scale items with the other half (Rust & Golombok 2009a; Langdridge & Hagger-Johnson 2013a; Coolican 2014). Some authors suggest the full test is administered and items split in half retrospectively to generate paired data (Coolican 2014), whereas others suggest prospectively splitting the items and administering different sets to 2 groups of people, generating unrelated data (Langdridge & Hagger-Johnson 2013a). Either way, correlated scores (or with an appropriate split half coefficient) would indicate high internal consistency (Rust & Golombok 2009a; Langdridge & Hagger-Johnson 2013a; Coolican 2014). To increase the relevance of this test to the whole scale, a Spearman–Brown formula can be applied (Rust & Golombok 2009a; Coolican 2014).
	Cronbach's alpha (α)	<p>The most popular method of testing scale internal consistency is the use of the Cronbach's α test, which relies on variance among items within each person's responses and is akin to averaging all possible split-half reliability values from the data set (Coolican 2014). Cronbach's $\alpha < 0.5$ is unacceptable, 0.5–0.6 is poor, 0.6–0.7 is questionable, 0.7–0.8 is acceptable, 0.8–0.9 is good, and > 0.9 is excellent, although there is scholarly debate in regards to their interpretation (Shelby 2011). Scores > 0.9 could suggest item redundancy (i.e., items are testing the same component but with slightly different wording), rather than providing a measure that contributes to the overall construct.</p> <p>Cronbach's α has been criticized for being biased, missing important properties when used in isolation, relying on often unmet assumptions, and providing false confidence in a scale. It may not be as powerful and robust as widely assumed (Zinbarg et al. 2005; Furr 2011a; Dunn et al. 2014). The almost universal use of this parameter in psychological research may be explained by a poor understanding of reliability analysis among researchers (Shelby 2011; Dunn et al. 2014). Reliability coefficients increase with increasing survey items, so researchers with large item sets should be more conservative in their interpretation of Cronbach's α, and correction for item-total correlations (see below) may be required (Furr 2011a; Shelby 2011). Similarly, no single measure of reliability should be used. Even with good Cronbach's α scores, a scale should be assessed for external consistency and item analysis conducted.</p>

Continued

Table 1. Continued.

Reliability measure	Test	Description
	Revelle's beta (β) and McDonald's omega (ω)	Alternatives to Cronbach's α exist, including the Revelle's β and MacDonald's ω (Zinbarg et al. 2005); ω has advantages, such as making fewer assumptions, reduced problems with inflation and attenuation, and better reflection of the variability in the estimation process (Dunn et al. 2014). Each of these measure different aspects of reliability, such as general factor saturation (extent to which all items measure the same construct), interrelatedness of items (distinct from homogeneity which refers to the unidimensionality of a set of items), or consistency of an assemblage of methods proposed to estimate single-administered test reliability (Tang et al. 2014). However, even when 6 indices of reliability were assessed (coefficient α , Cronbach's r_{ij} , Revelle's β , McDonald's ω , Sijsma's spell out, and SD), none were able to holistically or comprehensively measure internal consistency if used alone (Tang et al. 2014).
	parallel (or alternative) forms	Two sets of items are generated that are linked in a systematic manner and measure the same construct (Rust & Golombok 2009a). These 2 forms of the same scale are then pretested for correlation. Most often applied to knowledge tests, where 1 answer can be elicited from 2 versions of a question, this method is not typically suited to more abstract constructs, such as attitudes. However, it is relatively a common practice to use 2 forms of a scale during pretesting and to then select the best performing items and merge them into a final scale reflecting the best of both forms (Rust & Golombok 2009a). This may be appropriate for use in HWI attitude studies but should not necessarily be considered a measure of reliability when used in this manner.
	item-total correlation	Item-total correlation identifies items that are not consistent with the general trend and may, therefore, be irrelevant to the attitude of interest (Langdridge & Hagger-Johnson 2013a; Coolican 2014). An item's ability to discriminate between people scoring high and low overall can be used to refine the scale (Coolican 2014). Removal of nondiscriminating items must be checked for influence on construct validity before proceeding.
	item analysis	Item analysis is used to increase internal consistency. Analysis includes calculation of an item facility index (its mean score should be close to the center of the scale), assessment for frequency of response endorsement (items should not have >2 adjacent scale points with <20% of responses in total) and missing data (which should not exceed 4% for any item), and item discrimination (corrected item-total correlations should be >0.3). Following the removal of items not meeting these criteria, Cronbach's α coefficient is calculated. If 1 or more items are identified as reducing the overall consistency, these items can be removed from the scale (Langdridge & Hagger-Johnson 2013a; Coolican 2014). However, if items identified as contributing to poor internal consistency are deemed important for construct validity, their wording or structure should be scrutinized for ambiguity, irrelevance, or vagueness and then revised and retested rather than simply removing them at the first step (Coolican 2014).

form of negative HWI is known to affect a large number of predatory species (Seoraj-Pillai & Pillay 2017). Therefore, we specifically targeted this taxon (although not to the exclusion of other species).

We performed a literature search in the ProQuest PsychInfo and Web of Science databases. The terms “(wildlife or carnivore or predator) AND (attitude* or toleran*)” were searched for in the title of articles published from 2007 to 2017 (representing the decade immediately prior to this study being conducted). A total of 969 articles were returned and refined. Only peer-reviewed scholarly articles (i.e., excluding books and theses), written in English, and reporting first-generation investigations (i.e., excluding reviews or meta-analyses) were included. We screened article abstracts and excluded articles investigating knowledge assessment alone (e.g., Rutina et al. 2017), consumptive

use of wildlife, captive animals, human behavioral studies (unless attitudes were measured), media coverage, human-human attitudes, and domestic species (including feral domestic species). Only topics relevant to HWI in the context of conflict over wildlife or human coexistence with wildlife (i.e., excluding studies of other environmental, tourism, nature, or conservation topics) and related directly to the wildlife rather than management policies were included. Species-specific studies which did not include the generic terms (carnivore, predator, wildlife) in their titles were inadvertently excluded. Therefore, we acknowledge that this search does not reflect the entirety of HWI literature and is somewhat biased toward human-carnivore interactions for reasons outlined above. Likewise, extending the search to include all individual species involved in HWI was beyond the scope of this study. Our intention was to

Table 2. Measures of validity of psychometric scales relevant to human–wildlife interaction studies.

<i>Measure of validity</i>		<i>Description</i>
Face validity		Face validity exists when a scale in which the target attitude to be measured is apparent to both researchers and test takers (McCroskey 2006; Rust & Golombok 2009a; Langdridge & Hagger-Johnson 2013a; Coolican 2014). This test is not considered as robust as more quantitative methods of testing validity, but often provides an initial indication (McCroskey 2006) and must not be underestimated (Rust & Golombok 2009a). A scale with this validity has the advantage of potentially increasing motivation to participate (Rust & Golombok 2009a), along with the disadvantage of being easily falsified, but either way it is not an appropriate measure of validity for a scale when used in isolation (Langdridge & Hagger-Johnson 2013a; Coolican 2014). Control or placebo groups are useful for determining falsifiability (Coolican 2014), but this does not definitively demonstrate construct validity* because surveys may be either too broad or too narrow in scope to measure the attitude of interest comprehensively.
Construct validity*	content validity	Scales that include content that is irrelevant to the construct of interest omits relevant content, or underrepresents the true extent of content relevant to the construct is said to lack content validity (Furr 2011c). Content validity can be determined by a panel of experts or a literature review or both and, therefore, differs from face validity in the use of expert assessors (Furr 2011c). However, it still fails to provide a quantifiable or objective measure of validity (Langdridge & Hagger-Johnson 2013a; Coolican 2014).
	structural validity	The dimensionality of a scale must be appropriate to the construct of interest. Typically, simple or more easily defined constructs should be measurable with a unidimensional scale, but constructs with multiple dimensions are relatively common in human–wildlife interaction (HWI). The number of dimensions measured by a scale should reflect the theoretically expected number (Furr 2011c) (i.e., based on psychological theory pertaining to the construct). In the case of attitudes, it could be expected that affective, cognitive, or behavioral domains (or a combination of these) may exist. Determining the structural validity of a scale is performed using exploratory factor analysis (EFA), and subsequent confirmatory factor analysis. When an underlying trait is assumed to exist, EFA may be appropriate as this technique clusters items without accounting for error. However, when this assumption cannot be met, principal component analysis (PCA) is preferred because it incorporates error when clustering items and does not assume an underlying structure exists. Scree-plot analysis (a line plot of the Eigenvalues [a set of scalars associated with matrix equations] of the principal factors) is a superior means of determining the number of factors (or dimensions) inherent in a scale following EFA or PCA. However, the criterion commonly used is Eigenvalue > 1; this is likely to overestimate the number of factors and is the least accurate method of factor determination (Furr 2011c). When scree-plot analysis is inconclusive (e.g., there is no clear inflexion point), parallel analysis can be used as a robust method of factor extraction (Rust & Golombok 2009b; Furr 2011c). Parallel analysis functions by comparing the Eigenvalues of the factors identified in the data set with those calculated for a data set of random numbers (that have the same number of observations and variables as the real data). Factors with Eigenvalues in the real data that are greater than those generated from the random data are considered valid.
	convergent validity	Convergent validity measures score correlation with tests for related constructs (Rust & Golombok 2009a; Furr 2011c). The literature can be used to identify such related constructs and these should be tested alongside the new scale during pretesting. The method called multitrait, multimethod has been proposed as a more robust measure of construct validity following convergent validity testing (Rust & Golombok 2009a). This involves the testing of at least 3 traits with 3 different methods (generating a 9 × 9 correlation matrix) (Rust & Golombok 2009a).
	discriminant validity	Discriminant validity is effectively the opposite of convergent validity. A test with high discriminant validity should not correlate with measures of unrelated constructs (Rust & Golombok 2009a; Furr 2011c; Langdridge & Hagger-Johnson 2013a). By selecting constructs proven to lack correlation or alternatively lack a logical association (in the absence of previous studies to evidence this) with the construct of interest, it is possible to simultaneously measure both constructs in respondents during pretesting in order to test for discriminant validity. This form of validity is often underappreciated by researchers (Furr 2011c).

Continued

Table 2. Continued.

Measure of validity		Description
Criterion-related validity	concurrent validity	<p>Concurrent validity (also considered a component of construct validity) uses an existing measure of the same construct to compare with the new scale (Langdridge & Hagger-Johnson 2013a; Coolican 2014). This differs from convergent validity, which uses scales of related constructs. Using previously proven scales as a gold standard to compare newly developed scales is fundamental to validating a new scale (Rust & Golombok 2009a). Inherent within this measure is the assumption that previously validated scales exist and represent an accurate and reliable measure of the construct, which is unlikely to be entirely true. For this reason, concurrent validity should not be used in isolation, but if a new scale correlates well with an existing scale, then the researcher is provided with some confidence to pursue the new scale's validation (Rust & Golombok 2009a). Moreover, the sample-dependent nature of validation predicts that correlation between scales is unlikely to ever achieve perfect correlation. Counterintuitively, creation of a new scale when other scales already exist may be necessary (e.g., the investigation of a novel attitudinal object or component of attitudes to HWI, a novel context, or where existing scales are deemed inappropriate [e.g., divergent cultural, religious, or socioeconomic contexts] or flawed). Similarly, even where apparently appropriate scales exist, their psychometric parameters may not be reported, and in many cases they were not developed for comparative purposes or for use in other study contexts (Shah & Mahmood 2011). In these situations, such scales would not be considered gold standard and would not be appropriate for use in concurrent validity testing.</p> <p>Some criterion-based validity tests use a group for which the construct of interest is already known as a validation sample (Coolican 2014). Use of a validation group may be appropriate in cases of clinical psychology but is rarely (if ever) feasible when investigating attitudes toward HWI.</p>
	predictive validity	<p>This tests the ability of the scale to predict a criterion variable, which differs to the construct the scale is designed to measure (Langdridge & Hagger-Johnson 2013a). Although it is rare that HWI studies have the opportunity to prospectively measure attitudes before an HWI occurs, there would be scope for testing the predictive value of an attitude scale in a before-after test of an HWI intervention method. However, the variable outcomes of intervention methods and the diversity of attitudes among stakeholders would complicate the interpretation of this test of validity.</p> <p>Having said that, using scales aimed at determining value orientations or environmental attitudes to predict attitude toward particular HWI scenarios may be useful and has been applied in other wildlife contexts (Hartel et al. 2015). However, as illustrated by a study of hunting behavior and wildlife value orientation (including enjoyment of wildlife and wildlife rights) in which very poor correlation was determined (Hrubes et al. 2001), caution should be applied when assuming linear or predictable relationships between attitudes or underlying values and subsequent behavior.</p>

*There is variability in the way authors classify construct and criterion-related validity tests. Some suggest that construct validity is a test in its own right (Rust & Golombok 2009a), whereas others indicate construct validity comprises multiple components (Furr 2011c). For the purposes of explaining the various tests available, we used Furr's (2011b) definition.

understand methodological trends in studies of human attitudes and not to perform a meta-analysis of published findings (see Kansky et al. 2014).

A total of 114 published studies were identified as meeting our search criteria. Publication rate per year increased (6 published in 2007 vs. 14 published in 2017) (Fig. 2). Just over half of the qualifying studies (62 [54%]) included predators (species specific or as a group), 14 investigated a nonpredatory species, and 38 referred only to wildlife or animal conservation without specific species named in their title. A similar literature search for an earlier but partially overlapping period (1991–2014) that focused only on big cats revealed a similar number

of studies (63) to our predator-specific hits (Krafte Holland et al. 2018). Of all articles identified in our search, 94 (82%) were questionnaire based. The remaining 20 (18%) comprised methods such as participant observation, grounded theory, Q methodology, and unstructured interview approaches with qualitative data analyses (Fig. 2). This bias toward questionnaires seems to have remained relatively consistent (averaging 87% questionnaire based in the first half of the decade and 81% in the most recent 5 years). Several studies used >1 method; therefore, reported statistics for comparisons between methods were not mutually exclusive and did not always sum to 100%.

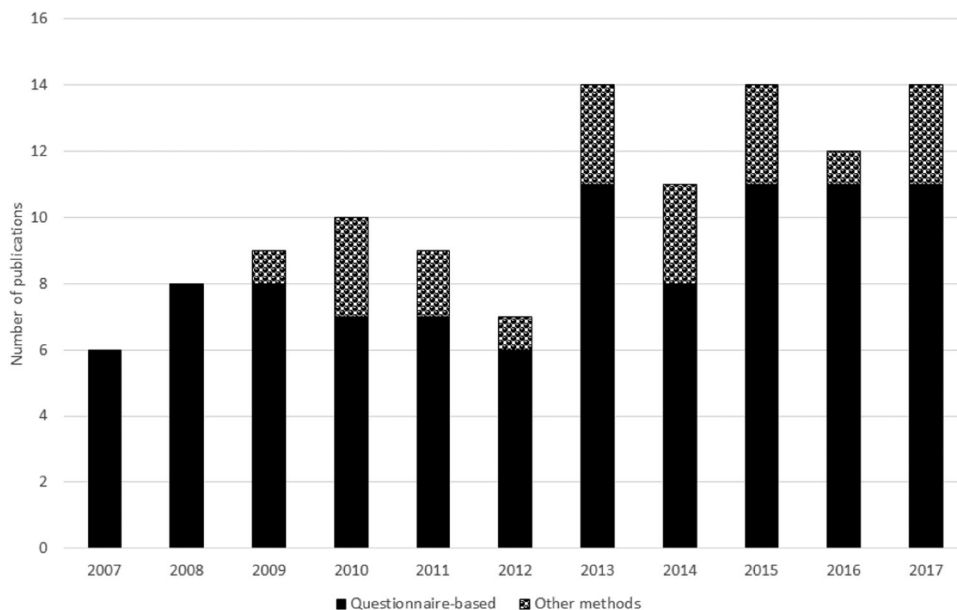


Figure 2. Number of articles in which the authors used questionnaire-based methods versus other methods within each year from 2007 to 2017. Articles were identified following a ProQuest and Web of Science literature search with the keywords attitude or tolerance in combination with wildlife or carnivore or predator; according to year of publication.

Questionnaires and Fixed-Structure Interviews

The most popular tool in HWI studies for collecting data from large groups was questionnaires. These were delivered either by self-administration (50 studies [53%]), or face-to-face or phone interview (44 studies; 47%) and are considered to offer the most structured method of data collection in the social sciences, allowing strict control over data collection and enabling deductive, quantitative approaches (Newing 2011a). Questionnaires (or fixed structure interviews) are considered objective and capable of detecting the presence or intensity of an attitude (or both), while producing generalizable results (Chase et al. 2016). Their frequent use in HWI studies may reflect the epistemology associated with the natural sciences, whereby conservation science has traditionally adopted quantitative approaches with little (if any) training in the social science methods (Moon & Blackman 2014; Sutherland et al. 2018; Martin 2020). Likewise, postal or online questionnaires may be popular due to their practical and logistical advantages (e.g., lower administration costs, access to larger samples in shorter periods of time, and reduced analytical time) over methods such as qualitative interviews (Myers et al. 2010).

Questionnaires have standardized set of questions generating targeted data that are relatively easy to analyze and make direct comparison among respondents (Newing 2011b). In HWI, a cross-sectional study aiming to generalize findings broadly may, therefore, benefit from the use of questionnaires. However, this structured approach may limit the scope or depth of responses achievable, risking a superficial evaluation of the attitude of interest (Chase et al. 2016). This is especially important when very specific problems or issues are the focus of investigation and the choice of quantitative or qualitative (or mixed) methods must reflect the research needs (Rust

et al. 2017). This is particularly relevant in cases where the attitude is interpreted as an indicator or proxy for behavior (Heberlein 2012; Nilsson et al. 2020). However, where longitudinal studies are concerned, the rigid structure of questionnaires offers the advantage of procedural standardization and allows direct comparison between 2 sampling points (Chase et al. 2016). Nonetheless, respondent attrition must be considered along with other factors (e.g., emotional, cognitive, behavioral or social experiences, and environments) that may influence attitude changes but could be overlooked when using fixed questions (Ajzen & Fishbein 1980; Nilsson et al. 2020).

Piloting of the questionnaire enables the researcher to determine whether the starting points for each group of stakeholder are equivalent in terms of their knowledge or attitude (Coolican 2014). The issue of nonequivalent groups is particularly important in HWI investigations evaluating conflict mitigation methods, education intervention schemes, or attempting to identify environmental or biological factors driving the HWI. Dissimilarities in attitudes between stakeholders may be equally (or more) important in determining the success or failure of an intervention than biological factors.

Pretesting of a small sample of people from the population of interest should investigate how respondents interpret the questions or statements, aiming to ensure the statements are not too complex and avoiding technical terms and ambiguity (Newing 2011a; Coolican 2014). In our assessment of HWI questionnaire studies, only 58 (51%) reported pretesting their instrument (this pretesting may have included studies that used previously established questionnaires for which pretesting had been reported during development or prior use). Some improvement was apparent in the most recent 5 years; 57% of recent studies reported using a pretest, compared with 40% in the first half of the decade.

Semistructured Interviews

Overall, 61 (54%) studies utilized some form of interview as a means of data capture, including either quantitative or qualitative analytical approaches (or both) to investigate attitudes or tolerance. In semistructured interviews, data can be acquired in an iterative process based on insights gained from conversational interviews to inform subsequent interviews (Ghosal & Kjosavik 2015). However, psychologists note some common nondirectional errors as a consequence of human cognitive biases and heuristics (e.g., Greenwald et al. 1995). Subsequently, interviews may be limited by inaccurate memory recall, lack of knowledge, generalization across events, reporting supposition rather than factual events, or distortion of memories by a respondent's prejudices (Newing 2011b).

Nonetheless, more personal interview styles can enable researchers to capture cultural, religious, legal, and moral forces at play in HWI (e.g., Ghosal & Kjosavik 2015). Under these scenarios, it is also possible for interviews to spontaneously evolve into focus groups, whereby associates of the interviewee can enter into the conversation (Ghosal & Kjosavik 2015). Focus groups were occasionally (4 studies, 4%) employed in HWI studies as part of the developmental phase for questionnaire development (e.g., Hazzah et al. 2009). Evaluation of focus groups and the other less commonly used methods in HWI studies is beyond the scope of this study, but methodological reviews for application in conservation more generally are available (e.g., Nyumba et al. 2018; Sutherland et al. 2018; Young et al. 2018).

Psychometric Scales in Current Practice

Of the questionnaire-based studies, just over half (53 studies, 56%) utilized a psychometric scale to determine participant attitudes toward wildlife. The use of scales has increased in the most recent 5 years; 38 studies reported their use compared with only 15 in the earlier 5 years.

Tests for reliability and validity are essential components to scale development. These tests (Tables 1 & 2) should be performed during pretesting and used to confirm the instrument suitability for the target population; they are then repeated for the final data set. This identifies the structural validity of the scale, from which construct-appropriate dimensionality can be confirmed. Overall, the majority of studies using such scales (39 studies, 74%) reported some form of reliability test, and 35 (66%) reported testing validity (prior publication was not considered sufficient because reuse in a new sample still requires a sample-specific validation step). Validity was typically restricted to factor

analysis (21 of 48 studies, 44%), which evaluates only the structural validity of a scale, whereas content validity, criterion-related validity, convergent, concurrent, or discriminant validity (Table 2) were rarely, if ever, reported. Similarly, reliability was primarily reliant on the Cronbach's alpha coefficient (33/44 studies, 75%). This statistic measures one aspect of internal consistency, which is an important component of scale reliability, but by no means all-encompassing (Table 1). External consistency was rarely reported, and other measures of internal consistency (e.g., item analysis and split-half reliability) were seldom used.

However, the reporting of validity increased within the most recent 5 years (71% of studies), compared with the first half of the decade (53% of studies). Contrastingly, testing for reliability (use of at least 1 test) decreased (71% most recently, compared with 80% earlier). None of the studies repeated the test to assess scale consistency with a test-retest strategy.

Bias in Measuring HWI-Relevant Attitudes

The primary source of bias in attitude measurements is response bias (or response set) (Table 3) (Langdridge & Hagger-Johnson 2013a; Coolican 2014), and efforts should be made to ensure nonresponse does not exceed 15% (Lindner et al. 2001; Lindner 2002). However, response rates to social surveys have declined over time (Gummer 2019); therefore, a response rate of 85% may now be unrealistic. Despite the decrease in response rate, it is encouraging that nonresponse bias appears relatively stable, at least in the populations investigated (Gummer 2019). Nonetheless, this reiterates the need to measure nonresponse rate to facilitate appropriate consideration of its effect during data interpretation.

Questionnaires are particularly vulnerable to nonresponse bias, but Dillman (1991) published a method for conducting mail surveys that has been adopted (in full or modified versions) for use in a number of HWI studies (e.g., Teel & Manfredi 2010; Thornton & Quinn 2010; Rodgers & Pienaar 2018) and that includes a discussion on ways to reduce nonresponse. Fourteen (15%) HWI studies utilizing questionnaires acknowledged nonrespondents. Of these, most (12 studies, 86%) attempted to determine associated bias. For example, in cases where the response rate is <85%, there are a number of methods available to handle this (Lindner et al. 2001; Lindner 2002), including follow-up interviews with nonresponders as used in a number of HWI studies. Alternatively, comparing the sample demographics of responders with an independent source of demographic data for the population of interest (e.g., by comparison with census data) may confirm sample representativeness (e.g., Vaske et al. 2011; Chase et al. 2016). However, uncontacted nonresponders may represent an

Table 3. Bias in studies measuring attitudes toward wildlife.

<i>Type of bias</i>	<i>Implications</i>	<i>Methods to reduce bias</i>
Response acquiescence	<p>The human tendency to agree with statements can introduce bias, whereby many people find it harder to provide negative responses or disagree with statements (Langdridge & Hagger-Johnson 2013a; Coolican 2014).</p> <p>This must be considered during study design because it can generate misleading analytical outcomes (Podsakoff et al. 2003), and survey instruments should be created in a way so as to avoid or minimize this risk.</p>	<p>In societies where norms may distort reporting of attitudes to sensitive topics, forced-choice questions may be preferable over continuum scales to reduce response acquiescence (Nuno & St. John 2014).</p> <p>A balance between negative and positive statements can assist in overcoming this source of bias.</p>
Social desirability bias	<p>This bias occurs when respondents provide answers they believe are most socially acceptable or attempt to answer in a way they assume the researcher is hoping or expecting (i.e., demand characteristics requiring impression management) (Podsakoff et al. 2003; Langdridge & Hagger-Johnson 2013a; Coolican 2014).</p>	<p>This bias may be especially prominent when data are being collected in person (including phone interviews), rather than anonymously (e.g., self-administered mail or online questionnaires). To identify this type of bias, some researchers include a statement or question in their survey instrument that provides an indication of the degree to which a respondent is misrepresenting themselves (Coolican 2014). These items (also called lie scales) could relate to a behavior that the majority of people would not perform in the extreme, whereby a <i>strongly agree</i> or <i>always</i> response to the statement could be indicative of social desirability bias (Langdridge & Hagger-Johnson 2013a). Social desirability scales also exist, and can be used in conjunction with the scale of interest.</p>
Leniency bias	<p>This may occur during data interpretation or coding (i.e., if researchers subconsciously or inadvertently alter their coding of respondents with whom they have an existing connection or have developed a relationship or attitude toward) (Podsakoff et al. 2003).</p>	<p>The use of multiple researchers and tests for interobserver or rater variability can be used to determine this type of bias.</p>
Awareness or context bias	<p>A heightened awareness of the topic may mean respondents' attitudes are highly context specific, reducing the potential for extrapolation of findings to other scenarios. This issue has been demonstrated for a widely used environmental attitude scale (New Ecological Paradigm scale) (Pienaar et al. 2013). Even previously uninformed respondents may be influenced by the statements used in the attitude scales, whereby items may require them to consider information or perspectives they had not considered previously (Manfredo & Bright 2008).</p> <p>Some stakeholders may conduct their own research prior to providing a response (if responses are gathered remotely rather than instantaneously without prior notice), or may even begin to analyze their own responses and adjust them in an attempt to better accentuate their position in the overall findings.</p>	<p>Effectively, the survey instrument itself could be influencing responses (Manfredo & Bright 2008), even when care is taken to avoid leading questions.</p> <p>However, experimental realism will be high for stakeholders involved in a human-wildlife interaction (HWI); therefore, it could be expected that their interest in the outcomes of the study will prompt attentiveness and a willingness to answer truthfully so that their perspective contributes to the outcome. In many HWI cases, replication or extrapolation to larger populations is not necessary or expected; therefore, this context specificity may be advantageous. Understanding the subjective nature of each data collection encounter is often a key goal in HWI research.</p>

Continued

Table 3. Continued.

<i>Type of bias</i>	<i>Implications</i>	<i>Methods to reduce bias</i>
Sensitive question bias	<p>It may be challenging to overcome social desirability bias because most HWI survey instruments are employed to measure attitudes in stakeholders who are actively involved in often emotionally complex, or even illegal or highly controversial HWI situations.</p> <p>Contrastingly, assumptions regarding the researchers' own attitude and intentions may provoke a defensive or even hostile stance in respondents. Likewise, respondents may adopt an antagonistic approach to the study (responding in a manner they predict to be contradictory to the desired or anticipated response) if they believe the perspective of the researcher is in opposition to their own interests (Coolican 2014).</p> <p>However, stakeholders have an interest in the topic and some may, therefore, invest additional effort in providing responses that they believe will further their cause.</p>	<p>Researcher independence is likely to be important in promoting honest and valid responses for sensitive topics in HWI, such that assurance of research impartiality and an objective and nonjudgmental interview environment will be critical (Langdridge & Hagger-Johnson 2013b; Nuno & St. John 2014). However, acknowledgment of researchers' opinions can also be important and could increase trust and openness (see above social desirability bias). Nonetheless, assurance of anonymization and the avoidance of highly personal or any unnecessary invasion of privacy will also be important (Coolican 2014; Nuno & St. John 2014).</p> <p>The use of specialized questioning techniques, such as randomized response technique, when asking sensitive questions in conservation can improve response rate and reduce social desirability bias (St. John et al. 2012; Nuno & St. John 2014).</p> <p>Consideration for the cultural, socioeconomic, and individual historical involvement in the HWI will be vital in establishing rapport.</p>
Biased assimilation	<p>Attitude polarization (i.e., respondents become more extreme in their views) can occur after respondents are provided with additional information regarding the attitudinal object (Miller et al. 1993).</p> <p>Similarly, when employing a repeated-measures study design, in which respondents' attitudes are measured prior to and then again following some form of intervention, the participation in the baseline survey may serve to heighten respondents' awareness of the topic. This could alter the degree of reflection or personal investment in knowledge acquisition during the interim period prior to the final attitude test and hence potentially affect their responses.</p>	<p>As per awareness and context bias.</p>
Affectivity or transient mood bias	<p>The affective (emotional) state of a person can influence the manner in which they respond to surveys (Podsakoff et al. 2003).</p>	<p>Attitudes are considered a relatively enduring phenomenon over the short term (relative to a lifetime). Therefore, attitude tests should be performed twice over a time interval to ensure views expressed are not simply transient opinions, rather than true attitudes (Coolican 2014). This differs to test-retest reliability, which is performed with a shorter interval and during the developmental stages of questionnaire design.</p> <p>Repeat testing is rarely (if ever) implemented except when attitude change is the focus of investigation and, therefore, does not fulfill the function of determining affectivity bias.</p>

Continued

Table 3. Continued.

<i>Type of bias</i>	<i>Implications</i>	<i>Methods to reduce bias</i>
Consistency motif effects, implicit theories, and illusory correlations	Consistency motif effects occur when people attempt to maintain a consistent approach to questions, rather than answering each independently (Podsakoff et al. 2003). Alternatively, respondents can introduce implicit theories and illusory correlations (perceiving a relationship when none exists) when assumptions are made regarding predictor and criterion variables (Podsakoff et al. 2003).	This is particularly problematic when asking respondents to retrospectively define their attitudes (Podsakoff et al. 2003). Response analysis and interview-style pretesting of questions or statements may identify these effects.

important but unmeasured section of society, potentially with divergent views on the attitudinal object to the responders, that warrant careful consideration.

Other forms of bias also exist, and these require consideration during study design and data interpretation (Table 3). For example, demand characteristics are an artifact of surveys in which respondents provide answers aligned to what they believe the researcher may expect (or demand), rather than their own opinion. These are important confounders, and researchers must consider how best to eliminate cues, which may convey researcher expectancy or otherwise influence respondents reactions (Coolican 2014). These cues may be unavoidable if the researcher is working for or with groups involved in the HWI, but could also be conveyed inadvertently in the wording of the study aims or via interpersonal communication (Coolican 2014). Calls for researchers to acknowledge their own stakeholder role in HWI, and conservation generally, suggest a more open acknowledgment of their standpoints should be encouraged (Redpath et al. 2013; Hill 2015; Rust & Taylor 2016).

Conclusions

A strong preference for quantitative methods exists within the HWI literature when measuring human attitudes. Although this in itself is not necessarily problematic, concern regarding inappropriate application of social science methods to address environmental issues is justified (Heberlein 1988, 2012; Moon & Blackman 2014; Martin 2020), at least when it comes to studies of attitudes toward wildlife. Most notably, our results reveal a limited application of appropriate reliability and validity tests. This undermines the capacity of many studies to derive any useful or legitimate information regarding attitudes or tolerance toward wildlife.

Although the use of a prevalidated scale may be an attractive option for interpopulation comparisons, the validity of an attitude measure cannot be assumed to remain constant when applied in a different context or culture to which the scale was first devised and tested

(Paunonen & Ashton 1998). Extrapolation to other people (population validity), settings (ecological validity), or periods of time (historical validity) is often a focal point of discussions arising from studies of attitudes toward wildlife, but researchers should be cautious of deriving inferential data from samples when dealing with complex constructs such as attitudes. Nonetheless, publication of more comprehensively tested scales creates opportunity for comparisons between studies and generalizability of findings. Likewise, the availability of a range of validated and reliable scales for HWI attitude constructs could assist in reducing the creation and use of poorly constructed or unreliable and unvalidated scales. Both scenarios carry the necessary prerequisite for determining population and ecological validity of existing scales prior to use.

For inter or cross-disciplinary HWI studies to achieve their full potential contribution to conservation, equivalent attention must be paid to best practice in study design and instrumentation for social data collection as is paid to biological data collection and analyses. This necessitates engagement with the large body of literature available to support psychological research. We provided a summary analysis of some of this literature, highlighting the key factors involved in just one psychological research method. Our tables were devised to achieve our aim and reflect the complexity of the topic and the necessary breadth and depth of understanding associated with the use of psychometric scales. Entire libraries would be required to incorporate the multitude of other research methods of relevance to HWI. Moreover, we rely heavily on other sources to refer readers to broader aspects of the psychological theories of attitudes, and we have not begun to cover conservation psychology as a whole. Therefore, we join previous authors (e.g., Heberlein 1988; Montgomery et al. 2018; Martin 2020) in advocating for more truly integrated inter and cross-disciplinary research, including collaborations with trained, experienced social scientists. This approach is required in combination with greater awareness, appreciation, and understanding of social science methods by conservation biologists. We

believe this tactic will facilitate improved practice in the application of psychological research methods in HWI research and maximize the opportunities for scientific advancement in conservation.

Acknowledgments

Nottingham Trent University Quality Research funding supported K.W-T. during the writing of this paper.

Supporting Information

A summary of the psychometric scales commonly used in HWI (as shown in Fig. 1), including a description of their construction and key considerations for their use (Appendix S1) is available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

- Ajzen I. 1985. From intentions to actions: a theory of planned behavior. Pages 11–39 in Kuhland J, Beckman J, editors. *Action-control: from cognitions to behavior*. Springer, Heidelberg.
- Ajzen I, Fishbein M. 1980. *Understanding attitudes and predicting social behavior*. Prentice-Hall, Englewood-Cliffs, New Jersey.
- Ajzen I, Timko C, White JB. 1982. Self-monitoring and the attitude-behavior relation. *Journal of Personality and Social Psychology* **42**:426–435.
- Baker PJ, Boitani L, Harris S, Saunders G, White PCL. 2008. Terrestrial carnivores and human food production: impact and management. *Mammal Review* **38**:123–166.
- Chase LD. 2016. Measurement of wildlife value orientations among diverse audiences: a multigroup confirmatory factor analysis among Hispanic and non-Hispanic white communities. *Human Dimensions of Wildlife* **21**:127–143.
- Chase LD, Teel TL, Thornton-Chase MR, Manfredo MJ. 2016. A comparison of quantitative and qualitative methods to measure wildlife value orientations among diverse audiences: a case study of Latinos in the American Southwest. *Society and Natural Resources* **29**:572–587.
- Coolican H. 2014. Psychological tests and measurement scales. Pages 193–229 in Coolican H, editor. *Research methods and statistics in psychology*. Routledge, London.
- Decker DJ, Riley SJ, Siemer WF. 2012. *Human dimensions of wildlife management*. John Hopkins University Press, Baltimore, Maryland.
- Delibes-Mateos M. 2014. Negative attitudes toward predators do not necessarily result in their killing. *Oryx* **48**:16.
- Dickman AJ. 2008. Key determinants of conflict between people and wildlife, particularly large carnivores, around Ruaha National Park, Tanzania. PhD thesis. University College London & Zoological Society of London, London.
- Dickman AJ. 2010. Complexities of conflict: the importance of considering social factors for effectively resolving human-wildlife conflict. *Animal Conservation* **13**:458–466.
- Dietsch AM, Manfredo MJ, Teel TL. 2017. Wildlife value orientations as an approach to understanding the social context of human-wildlife conflict. Page 228 in Hill CM, Webber AD, Priston NEC, editors. *Understanding conflicts about wildlife: a biosocial approach*. Berghahn, New York.
- Dillman Don A. 1991. The design and administration of mail surveys. *Annual Review of Sociology* **17**:225–249. <http://doi.org/10.1146/annurev.so.17.080191.001301>.
- Dunn TJ, Baguley T, Brunsden V. 2014. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology* **105**:399–412.
- Eagly AH, Chaiken S. 1993. *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers, San Diego, California.
- Eagly AH, Mladinic A, Otto S. 1994. Cognitive and affective bases of attitudes toward social groups and social policies. *Journal of Experimental Social Psychology* **30**:113–137.
- Fischer R, Milfont T. 2010. Standardization in psychological research. *International Journal of Psychological Research* **3**:88–96.
- Fulton DC, Manfredo MJ, Lipscomb J. 1996. Wildlife value orientations: a conceptual and measurement approach. *Human Dimensions of Wildlife* **1**:24–47.
- Furr RM. 2011a. Evaluating psychometric properties: dimensionality and reliability. Pages 25–51 in Furr RM, editor. *Scale construction and psychometrics for social and personality psychology*. Sage Publications, London.
- Furr RM. 2011b. Core principles, best practices, and an overview of scale construction. Pages 4–15 in Furr RM, editor. *Scale construction and psychometrics for social and personality psychology*. Sage Publications, London.
- Furr RM. 2011c. Evaluating psychometric properties: validity. Pages 52–66 in Furr RM, editor. *Scale construction and psychometrics for social and personality psychology*. Sage Publications, London.
- Ghosal S, Kjosavik DJ. 2015. Living with leopards: negotiating morality and modernity in western India. *Society and Natural Resources* **28**:1092–1107.
- Greenwald AG, et al. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review* **102**:4–27.
- Gummer T. 2019. Assessing trends and decomposing change in non-response bias: the case of bias in cohort distributions. *Sociological Methods and Research* **48**:92–115.
- Hartel CM, Carlton JS, Prokopy LS. 2015. The role of value orientations and experience on attitudes toward a well-liked threatened reptile. *Human Dimensions of Wildlife* **20**:553–562.
- Hazzah L, Borgerhoff Mulder M, Frank L. 2009. Lions and warriors: social factors underlying declining African lion populations and the effect of incentive-based management in Kenya. *Biological Conservation* **142**:2428–2437.
- Heberlein TA. 1988. Improving interdisciplinary research: integrating the social and natural sciences. *Society and Natural Resources* **1**:5–16.
- Heberlein TA. 2012. Navigating environmental attitudes. *Conservation Biology* **26**:583–585.
- Hill CM. 2015. Perspectives of “conflict” at the wildlife-agriculture boundary: 10 years on. *Human Dimensions of Wildlife* **20**:296–301.
- Hogg M, Vaughan G. 2005. *Social psychology*. 4th edition. Prentice-Hall, London.
- Hrubes D, Ajzen I, Daigle J. 2001. Predicting hunting intentions and behavior: an application of the theory of planned behavior. *Leisure Sciences* **23**:165–178.
- Kansky R, Kidd M, Knight AT. 2014. Meta-analysis of attitudes toward damage-causing mammalian wildlife. *Conservation Biology* **28**:924–938.
- Kansky R, Knight AT. 2014. Key factors driving attitudes toward large mammals in conflict with humans. *Biological Conservation* **179**:93–105.
- Katz D. 1960. The functional approach to the study of attitudes. *Public Opinion Quarterly* **24**:163–204.
- Krafte Holland K, Larson LR, Powell RB. 2018. Characterizing conflict between humans and big cats *Panthera* spp: a systematic review of research trends and management opportunities. *PLOS ONE* **13**:e0203877. <http://doi.org/10.1371/journal.pone.0203877>.

- Langdridge D, Hagger-Johnson G. 2013a. Collecting data: questionnaires and psychometric tests. Pages 87–112 in Langdridge D, Hagger-Johnson G, editors. *Introduction to research methods and data analysis in psychology*. Pearson Education, London.
- Langdridge D, Hagger-Johnson G. 2013b. Collecting data: interviews and observations. Pages 61–86 in Langdridge D, Hagger-Johnson G, editors. *Introduction to research methods and data analysis in psychology*. Pearson Education, London.
- LaPiere RT. 1934. Attitudes vs. actions. *Social Forces* **13**:230–237.
- Lindner JR. 2002. Handling of nonresponse error in the *Journal of International Agricultural and Extension Education*. *Journal of International Agricultural and Extension Education* **9**:55–60.
- Lindner JR, Murphy TH, Briers GE. 2001. Handling nonresponse in social science research. *Journal of Agricultural Education* **42**:43–53.
- Maddox TM. 2003. The ecology of cheetahs and other large carnivores in a pastoralist-dominated buffer zone. PhD dissertation. University College London & Zoological Society London, London.
- Manfredo MJ, Bright AD. 2008. Attitudes and the study of human dimensions of wildlife. Pages 75–110 in Manfredo MJ, editor. *Who cares about wildlife?* Springer, New York.
- Martin VY. 2020. Four common problems in environmental social research undertaken by natural scientists. *BioScience* **70**:13–16.
- McCroskey JC. 2006. Reliability and validity of the generalized attitude measure and generalized belief measure. *Communication Quarterly* **54**:265–274.
- Miller AG, McHoskey JW, Bane CM, Dowd TG. 1993. The attitude polarization phenomenon: role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology* **64**:561–574.
- Montgomery RA, et al. 2018. Examining evident interdisciplinarity among prides of lion researchers. *Frontiers in Ecology and Evolution* **6**:1–13.
- Moon K, Blackman D. 2014. A guide to understanding social science research for natural scientists. *Conservation Biology* **28**:1167–1177.
- Myers D, Abell J, Sani F. 2010. Attitudes and behavior. Pages 157–200 in Myers D, Abell J, Kolstad A, Sani F, editors. *Social psychology*. McGraw-Hill Education, London.
- Newing H. 2011a. Questionnaires. Pages 119–145 in Newing H, editor. *Conducting research in conservation: social science methods and practice*. Routledge, Oxon.
- Newing H. 2011b. Qualitative interviews and focus groups. Pages 98–118 in Newing H, editor. *Conducting research in conservation: social science methods and practice*. Routledge, Oxon.
- Nilsson D, Fielding K, Dean A. 2020. Achieving conservation impact by shifting focus from human attitudes to behaviors. *Conservation Biology* **34**:93–102.
- Nuno A, St. John FAV. 2014. How to ask sensitive questions in conservation: a review of specialized questioning techniques. *Biological Conservation* **189**:5–15.
- Nyumba TO, Wilson K, Derrick CJ, Mukherjee N. 2018. The use of focus group discussion methodology: insights from two decades of application in conservation. *Methods in Ecology and Evolution* **9**:20–32.
- Paunonen S, Ashton MC. 1998. The structured assessment of personality across cultures. *Journal of Cross-cultural Psychology* **29**:150–170.
- Perry L, Loveridge A, Macdonald D, Moorhouse T. 2020. The role of psychology in determining human–predator conflict across southern Kenya. *Conservation Biology* **34**:879–890.
- Pienaar EF, Lew DK, Wallmo K. 2013. Are environmental attitudes influenced by survey context? An investigation of the context dependency of the New Ecological Paradigm (NEP) scale. *Social Science Research* **42**:1542–1554.
- Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* **88**:879–903.
- Pooley S, et al. 2017. An interdisciplinary review of current and future approaches to improving human–predator relations. *Conservation Biology* **31**:513–523.
- Redpath SM, et al. 2013. Understanding and managing conservation conflicts. *Trends in Ecology and Evolution* **28**:100–109.
- Rodgers PD, Pienaar EF. 2018. Tolerance for the Florida panther in ex-urban southwest Florida. *Journal of Wildlife Management* **82**:865–876.
- Rust J, Golombok S. 2009a. Characteristics of tests. Pages 71–88 in Rust J, Golombok S, editors. *Modern psychometrics: the science of psychological assessment*. Routledge, Sussex.
- Rust J, Golombok S. 2009b. Factor analysis. Pages 105–120 in Rust J, Golombok S, editors. *Modern psychometrics: the science of psychological assessment*. Routledge, Sussex.
- Rust NA, et al. 2017. Quantity does not always mean quality: the importance of qualitative social science in conservation research. *Society & Natural Resources* **30**:1–7.
- Rust NA, Taylor N. 2016. Carnivores, colonization, and conflict: a qualitative case study on the intersectional persecution of predators and people in Namibia. *Anthrozoos* **29**:653–667.
- Rutina LP, Mogwera KM, Seonyatseng E, Mpofu C, Ntloyathuto D. 2017. Herders' ecological knowledge and carnivore predation on livestock investigations in Makgadikgadi and Nxai national parks, Botswana. *Koedoe* **59**:1–9.
- Seoraj-Pillai N, Pillay N. 2017. A meta-analysis of human–wildlife conflict: South African and global perspectives. *Sustainability* **9**:1–21.
- Shah ZA, Mahmood N. 2011. Developing a scale to measure attitude toward science learning among school students **33**:71–81.
- Shelby LB. 2011. Beyond Cronbach's alpha: considering confirmatory factor analysis and segmentation. *Human Dimensions of Wildlife* **16**:142–148.
- St. John FAV, Keane AM, Edwards-Jones G, Jones L, Yarnell RW, Jones JPG. 2012. Identifying indicators of illegal behaviour: carnivore killing in human-managed landscapes. *Proceedings of the Royal Society B: Biological Sciences* **279**:804–812.
- Stevens SS, Amulike B, Ndaga S, Organ JF, Serfass TL. 2014. The confusion of common names: a methodological challenge. *Human Dimensions of Wildlife* **19**:191–199.
- Sutherland WJ, Dicks LV, Everard M, Geneletti D. 2018. Qualitative methods for ecologists and conservation scientists. *Methods in Ecology and Evolution* **9**:7–9.
- Tang W, Cui Y, Babenko O. 2014. Internal consistency: do we really know what it is and how to assess it? *Journal of Psychology and Behavioral Science* **2**:205–220.
- Teel TL, Manfredo MJ. 2010. Understanding the diversity of public interests in wildlife conservation. *Conservation Biology* **24**:128–139.
- Thornton C, Quinn MS. 2010. Risk perceptions and attitudes toward cougars in the southern foothills of Alberta. *Human Dimensions of Wildlife* **15**:359–372.
- Treves A, Bruskotter J. 2014. Tolerance for predatory wildlife. *Science* **344**:476–477.
- Vaske JJ, Donnelly MP. 1999. A value-attitude-behavior model predicting wildland preservation voting intentions. *Society and Natural Resources* **12**:523–537.
- Vaske JJ, Jacobs MH, Sijsma MTJ. 2011. Wildlife value orientations and demographics in The Netherlands. *European Journal of Wildlife Research* **57**:1179–1187.
- Vaske JJ, Whittaker D. 2004. Normative approaches to natural resources. Pages 283–294 in Manfredo MJ, Vaske JJ, Bruyere BL, Field DR, Brown P, editors. *Society and natural resources: a summary of knowledge*. Modern Litho, Jefferson, Missouri.
- Young JC, et al. 2018. A methodological guide to using and reporting on interviews in conservation science research. *Methods in Ecology and Evolution* **9**:10–19.
- Zinbarg R, Revelle W, Yovel I, Li W. 2005. Cronbach's alpha, Revelle's beta, and McDonald's omega: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* **70**:123–133.